
Characterizing Regularization Techniques for Spatial Filter Optimization in Oscillatory EEG Regression Problems

Guidelines Derived from Simulation and Real-World Data

Andreas Meinel · Sebastian
Castaño-Candamil · Benjamin Blankertz ·
Fabien Lotte · Michael Tangermann

Received: 02.02.2018 / Accepted: 31.07.2018

Abstract We report on novel supervised algorithms for single-trial brain state decoding. Their reliability and robustness are essential to efficiently perform neurotechnological applications in closed-loop. When brain activity is assessed by multichannel recordings, spatial filters computed by the source power comodulation (SPoC) algorithm allow identifying oscillatory subspaces. They regress to a known continuous trial-wise variable reflecting, e.g. stimulus characteristics, cognitive processing or behavior. In small dataset scenarios, this supervised method tends to overfit to its training data as the involved recordings via electroencephalogram (EEG), magnetoencephalogram or local field potentials generally provide a low signal-to-noise ratio. To improve upon this, we propose and characterize three types of regularization techniques for SPoC: approaches using Tikhonov regularization (which requires model selection via cross-validation), combinations of Tikhonov regularization and covariance matrix normalization as well as strategies exploiting analytical covariance matrix shrinkage. All proposed techniques were evaluated both in a novel simulation framework and on real-world data. Based on the simulation findings, we saw our expectations fulfilled, that SPoC regularization generally reveals the largest benefit for small training sets and under severe label noise conditions. Relevant for practitioners, we derived operating ranges of regularization hyperparameters for cross-validation based approaches and offer open source code. Evaluating all methods additionally on real-world data, we observed an improved regression performance mainly for datasets from subjects with initially poor performance. With this proof-of-concept paper, we provided a gen-

A. Meinel · S. Castaño-C. · M. Tangermann
Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools, Dept. of Computer Science, Albert-Ludwigs-University, Freiburg, Germany
E-mail: {andreas.meinel},{michael.tangermann}@blbt.uni-freiburg.de

B. Blankertz
Neurotechnology Group, Technische Universität Berlin, Berlin, Germany

F. Lotte
Inria, Potioc Project Team, Talence, France
LaBRI (University of Bordeaux, CNRS, INP), Talence, France
RIKEN Brain Science Institute, Japan

eralizable regularization framework for SPoC which may serve as a starting point for implementing advanced techniques in the future.

Keywords EEG bandpower · subspace decomposition · brain-computer interface · single-trial analysis · source power comodulation · brain state decoding algorithm

1 Introduction

In many modern applications based on biomedical signals, machine learning software is extensively used to infer variables or states of interest Mahmud et al (2018). Examples are electrocardiographic imaging, where machine learning can be used to deduce cardiac activities or pathologies from multiple sensors (Ramanathan et al, 2004), for controlling an upper-limb prosthesis from an amputee’s electromyography recordings (Farina et al, 2014), or for decoding users’ mental states from their electroencephalographic (EEG) activity (Dähne et al, 2014; Clerc et al, 2016). In all these applications, signals are recorded using multiple sensors, resulting in multivariate data that should be analyzed using robust classification or regression methods. Such machine learning problems are often very challenging.

Prominent neurotechnological systems are brain-computer interfaces (BCI) which typically utilize EEG recordings that enable users to interact with a computer or physical devices (Millán et al, 2010). Such practical closed-loop applications require the extraction of relevant and robust features (Farquhar and Hill, 2013) from high-dimensional EEG data which unfortunately suffer from an inherently low signal-to-noise ratio (Krusienski et al, 2011; Makeig et al, 2012). In addition, for most BCI applications only small calibration datasets are available to train the decoding algorithms — typically a few dozens or maximally a couple of hundreds of training samples — which further aggravates the situation (Lotte, 2015). Thus, it is necessary to design robust decoding methods and training procedures, such that over-fitting to the training data is avoided (Makeig et al, 2012).

A widely used approach for effective decoding of EEG signals are spatial filter methods. They learn a linear transformation to project multivariate EEG signals derived from several sensors to a lower dimensional subspace (de Cheveigné and Parra, 2014; Blankertz et al, 2008) for instance to remove artifacts (De Vos et al, 2010) or to extract task-related neural activity (Makeig et al, 2004). In the context of BCI, the most prominent algorithm for a supervised scenario is the common spatial pattern algorithm (CSP; (Koles, 1991; Ramoser et al, 2000)). It is deployed for solving EEG classification tasks that are characterized by amplitude modulations of brain rhythms. Unfortunately, CSP is specifically sensitive towards noisy training data (Reuderink and Poel, 2008), non-stationarities (Samek et al, 2014) and small datasets (Grosse-Wentrup et al, 2009; Park et al, 2017). To mitigate a subset of these limitations, regularization variants have been proposed for CSP (Lotte and Guan, 2011; Samek et al, 2014). In general, regularization guides an optimization problem by adding prior information, thus limiting the space of possible solutions. Even though regularization is of specific importance for ill-posed problems such as source reconstruction (Tian et al, 2013), less underdetermined problems can also profit. For CSP, a broad bandwidth of regularization approaches has been published, such as L1- and L2-norm penalties (Wang and Li, 2016; Lotte

and Guan, 2011; Arvaneh et al, 2011; Farquhar et al, 2006), regularized transfer learning strategies that accumulate information across multiple sessions and subjects (Cheng et al, 2017; Devlaminck et al, 2011; Samek et al, 2013; Kang et al, 2009; Lotte and Guan, 2010) and variants which favor invariant solutions across sessions/runs under EEG non-stationarities (Arvaneh et al, 2013; Samek et al, 2012, 2014; Cho et al, 2015).

Taking a closer look into the BCI decoding literature, a variety of methods for oscillatory EEG classification problems can be found, but for the regression case the choice still is extremely limited (Wu et al, 2017) even though regression methods allow tackling highly interesting problems. Examples are the estimation of continuous mental workload levels (Frey et al, 2016; Schultze-Kraft et al, 2016), decoding the depth of cognitive processing (Nicolae et al, 2017), predicting single-trial motor performance (Meinel et al, 2016) or continuous decoding of movement trajectories (Úbeda et al, 2017). A spatial filtering solution, which solves an EEG regression problem, was provided by Dähne et al (2014) with the source power co-modulation algorithm (SPoC). It optimizes spatial filters that describe oscillatory subspace components, whose bandpower co-modulate with a given continuous univariate target variable. Comparing the formulations of the objective functions of SPoC and CSP, both can be translated into a Rayleigh quotient. Thus, the known limitations of CSP regarding noise, non-stationarities and limited data were found to also apply to SPoC (Castaño-Candamil et al, 2015). However, regularization approaches for SPoC have not yet been explored.

The main goal of this paper is to close this gap. Thus, we present generally applicable regularization variants for SPoC to improve the algorithm’s robustness. Therefore, we first evaluate regularized SPoC variants on a very recent simulation approach based on post-hoc labeling of arbitrary EEG recordings. This allows probing the stability of the regularized variants under reduced training datasets, varying label noise conditions and different strengths of oscillatory sources. In a second step, we evaluate the methods on a real-world dataset and compare the findings of both scenarios. As regularization introduces additional hyperparameters, we will compare model selection via cross-validation with an analytical solution. Finally, we provide the practitioner with information on how to determine suitable parameters for SPoC regularization and provide open source software for regularized SPoC. This paper builds upon preliminary results reported in a conference paper by Meinel et al (2017).

2 Supervised Spatial Filter Optimization for Single-Trial EEG Regression

Let $\mathbf{x}(t) \in \mathbb{R}^{N_c}$ describe multivariate EEG signals acquired from N_c channels at time sample t . Moreover, the signals are assumed to be bandpass filtered. A spatial filter $\mathbf{w} \in \mathbb{R}^{N_c}$ can be interpreted as a projection of $\mathbf{x}(t)$ from the N_c -dimensional sensor space to a one-dimensional source component $\hat{s}(t) = \mathbf{w}^\top \mathbf{x}(t)$ according to the linear model of the EEG Parra et al (2005).

2.1 Prerequisites for the SPoC algorithm

Based upon the course of events in the conducted experimental paradigm, an EEG recording $\mathbf{x}(t)$ can be translated into N_e single epochs $\mathbf{X}(e) \in \mathbb{R}^{N_c \times N_s}$ with N_s sample points per epoch. Hereafter, $\langle \cdot \rangle$ refers to an average across epochs. For each epoch e , we assume to have access to a continuous epoch-wise target variable $z(e)$ which is required to be standardized to zero mean and unit variance. This variable z provides label information about the experimental paradigm and could represent e.g. stimulus intensity, behavioral responses or cognitive measures. Subsequently, our goal is to search for a source component \hat{s} within the full EEG signal, whose epoch-wise power $\Phi_{\hat{s}}(e)$ is linearly related to the target variable $z(e)$.

A spatial filter \mathbf{w} defines an estimated source \hat{s} when applied to the EEG signal. Due to the preceding bandpass filtering, its power can be assessed by its variance within the epoch e such that $\Phi_{\hat{s}}(e) = \text{Var}[\hat{s}(t)](e) = \text{Var}[\mathbf{w}^\top \mathbf{x}(t)](e)$.

The central principle of SPoC is to search an optimal spatial filter \mathbf{w}^* such that the epoch-wise power of the resulting estimated source component \hat{s} maximally co-modulates with the known target variable z . Once this optimal filter has been found, it can then be employed to estimate the target variable $z(e)$ from the bandpassed signal $\mathbf{X}(t, e)$ ¹ using the variance:

$$z_{est}(e) \approx \Phi_{\hat{s}}(e) = \text{Var}[\hat{s}(t)](e) = \mathbf{w}^\top \Sigma(e) \mathbf{w} \quad (1)$$

where $\Sigma(e) = (N_s - 1)^{-1} \mathbf{X}(e)^\top \mathbf{X}(e)$ denotes the epoch-wise spatial covariance matrix.

Striving to find a formulation of the overall optimization function requires two ingredients. The first one is expressed by the (Euclidean) mean of the power $\langle \Phi_{\hat{s}}(e) \rangle$ across epochs by:

$$\langle \Phi_{\hat{s}}(e) \rangle \stackrel{Eq. (1)}{=} \mathbf{w}^\top \langle \Sigma(e) \rangle \mathbf{w} = \mathbf{w}^\top \Sigma_{avg} \mathbf{w} \quad (2)$$

with $\Sigma_{avg} := \langle \Sigma(e) \rangle$ defining the averaged covariance matrix across all N_e epochs.

The second ingredient is expressed as the covariance between the epoch-wise power of the source component and the target variable z . One can show the following relation Dähne et al (2014):

$$J_1(\mathbf{w}) := \text{Cov}[\Phi_{\hat{s}}(e), z(e)] = \mathbf{w}^\top \Sigma_z \mathbf{w} \quad (3)$$

where $\Sigma_z := \langle \Sigma(e) z(e) \rangle$ defines the label-weighted covariance matrix averaged over epochs. While the original SPoC formulation comprised two different optimization strategies, we will restrict further analysis to SPoC _{λ} which optimizes covariance instead of correlation, but allows deriving a closed-form solution of the spatial filters.

2.2 Optimizing covariance - SPoC _{λ} algorithm

As the covariance is directly affected by the scaling of its arguments, it requires a constraint upon possible solutions. In SPoC _{λ} this is tackled by a filter norm

¹ To simplify the notation, epoched data $\mathbf{X}(t, e)$ will further on be written as a matrix $\mathbf{X}(e)$.

constraint $J_2(\mathbf{w}) := \text{Var}[\hat{s]}(e) = \mathbf{w}^\top \Sigma_{avg} \mathbf{w} = 1$ which translates into the following Rayleigh quotient:

$$J_\lambda(\mathbf{w}) = \frac{J_1}{J_2} = \frac{\mathbf{w}^\top \Sigma_z \mathbf{w}}{\mathbf{w}^\top \Sigma_{avg} \mathbf{w}} \quad (4)$$

The optimization task $\mathbf{w}^* = \arg\max_{\mathbf{w}} J_\lambda(\mathbf{w})$ can be transferred into a generalized eigenvalue problem (de Cheveigné and Parra, 2014) and thus delivers a closed-form solution. Overall, the approach returns a full set $\{\mathbf{w}_j^*\}_{j=1,\dots,N_c}$ of N_c spatial filters with j indexing the rank. It is determined in descending order of the eigenvalues and thereby according to the covariance. Throughout the remaining paper, the SPoC $_\lambda$ algorithm is used. It will be referred to by the term SPoC.

3 Regularization for regression based spatial filtering

In most BCI scenarios small training datasets of a high dimensionality are encountered (Makeig et al, 2012). In this setting, SPoC shows an impairing sensitivity and thus might be prone to overfit the training data (Castaño-Candamil et al, 2015). A common machine learning strategy in such situations is to add prior information and thus regularize the objective function of an algorithm.

Similar to the regularization strategies proposed by Lotte and Guan (2011) for CSP, there are two possible branches of regularization strategies for the SPoC algorithm: The first is to directly add prior information on the level of the objective function in Eq. 4. This leads to a restriction of the solution space of possible filters. The second one directly addresses the involved empirical covariance matrices which suffer from small training sets and noisy data. Poorly estimated covariance matrices will not characterize the intended neural activity well. Therefore, regularization on the level of covariance matrices intends to improve their estimation and thus enhance the spatial filtering optimization. In the following, we will propose two regularization approaches, one from each branch of strategies.

3.1 Additional penalty on the objective function

Introducing a regularization to the objective function of SPoC can be achieved by adding a penalty term $P(\mathbf{w})$ to the denominator of the Rayleigh quotient stated in Eq. 4:

$$\tilde{J}_\lambda(\mathbf{w}) = \frac{\mathbf{w}^\top \Sigma_z \mathbf{w}}{(1 - \alpha)\mathbf{w}^\top \Sigma_{avg} \mathbf{w} + \alpha P(\mathbf{w})} \quad (5)$$

where $\alpha \geq 0$ is the regularization parameter that modulates the strength of the penalty. In general, the term $P(\mathbf{w})$ penalizes solutions of \mathbf{w} that do not fulfill a specified prior. Thereby it increases the algorithm's robustness to outliers and small training sets.

In this paper, we select a simple quadratic penalty of the form:

$$P(\mathbf{w}) = \mathbf{w}^\top I \mathbf{w} = \|\mathbf{w}\|^2 \quad (6)$$

using the identity matrix $I \in \mathbb{R}^{N_c \times N_c}$. This penalty is known as *Tikhonov regularization* Tikhonov (1963) and has similarly been introduced for CSP (Lotte and Guan, 2011). As the penalty $P(\mathbf{w})$ scales with the spatial filter norm, solutions

\mathbf{w} with small weights are preferred. Regarding utmost regularization strength in Eq. 5 expressed by $\alpha = 1$, the Rayleigh quotient simplifies to the one of the principal component analysis (PCA, (De Bie et al, 2005)) meaning that a PCA on the z -weighted covariance matrix is computed. For the introduced Tikhonov regularization of SPoC, model selection wrt. α can be done via cross-validation (CV).

Trace normalization

SPoC $_{\lambda}$ optimizes covariance which is not scale-invariant. This drawback might be mitigated by the norm constraint, but to directly compensate for the relative scaling of the covariance matrices in Eq. 5, a normalization of *all* covariance elements by the trace $tr[\cdot]$ might also be a suitable strategy as already proposed for CSP (Ramoser et al, 2000; Lu et al, 2010):

$$\hat{\Sigma}(e) = \frac{\Sigma(e)}{tr[\Sigma(e)]}; \quad \hat{\Sigma}_{avg} = \frac{\Sigma_{avg}}{tr[\Sigma_{avg}]} \quad (7)$$

Here, we investigate the effect of applying trace norm to $\Sigma(e)$ and Σ_{avg} entering Eq. 5, but not upon Σ_z as its label-weighting shall be maintained.

Equivalence to covariance shrinkage

Inserting the given Tikhonov penalty $P(\mathbf{w})$ of Eq. 6 into the objective function in Eq. 5, enables to factorize the denominator to a shrinkage of the averaged covariance matrix Σ_{avg} towards the identity matrix $I \in \mathbb{R}^{N_c \times N_c}$:

$$\tilde{\Sigma}_{avg} = (1 - \alpha)\Sigma_{avg} + \alpha I \quad (8)$$

By that we have shown that substituting Σ_{avg} by the shrunked version $\tilde{\Sigma}_{avg}$ in the objective function of SPoC (see Eq. 4) is equivalent to the Tikhonov formulation stated in Eq. 5 and 6.

3.2 Regularization of covariance matrices

In parallel to the proposed Tikhonov regularization which builds upon a CV procedure for model selection, there are faster ways of determining a suitable regularization strength. We will focus on two strategies for covariance shrinkage which allow to use an analytic solution to determine the regularization parameter.

Automatic shrinkage of sample covariance matrices

When estimating a sample covariance matrix $S \in \mathbb{R}^{N_c \times N_c}$ based on N_{train} training data samples, there was a systematic bias reported in the setting of $N_c > N_{train}$: large eigenvalues get overestimated while small eigenvalues tend to be underestimated (Bai and Silverstein, 2009). The situation can be improved by shrinking the covariance matrix S towards the identity matrix I (Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005):

$$\tilde{S} = (1 - \alpha^*)S + \alpha^* \nu I \quad (9)$$

Under the assumption of i.i.d. data and thus in the absence of outliers, Ledoit & Wolf derived a closed-form solution for the optimal shrinkage parameter α^* and the optimal scaling parameter ν by minimizing the expected mean squared error. For the exact closed-form solution of α^* and ν , we refer the reader to Ledoit and Wolf (2004); Schäfer and Strimmer (2005); Bartz and Müller (2014). This closed-form solution holds the advantage of directly computing an estimate of α^* without cross-validation. Note that the additional scaling factor ν takes a similar role as the trace normalization introduced for Tikhonov regularization with the difference that it only takes diagonal terms into account.

Automatic shrinkage of averaged covariance matrix

As shown in section 3.1, the Tikhonov penalty introduced for the SPoC objective function can be rewritten as a shrinkage of the covariance matrix Σ_{avg} , which was gained by averaging across the epoch-wise covariances $\Sigma(e)$. Thus, one can directly apply the closed-form solution for α^* and ν , but it first requires to estimate the averaged covariance matrix as $\Sigma_{avg} = (N_s \cdot N_e - 1)^{-1} \mathbf{x}_{cat}^\top \mathbf{x}_{cat}$ using a concatenated data matrix of all N_e epochs, namely $\mathbf{x}_{cat} = [\mathbf{x}(1), \dots, \mathbf{x}(N_e)] \in \mathbb{R}^{N_c \times (N_s \cdot N_e)}$. To compensate for signal non-stationarities, each data epoch $\mathbf{X}(e)$ should be corrected to channel-wise zero mean prior to concatenation.

Automatic shrinkage of epoch-wise covariance matrix

SPoC includes the label-weighted covariance matrix Σ_z which holds all the available label information. A direct covariance shrinkage for Σ_z was tested in pilot experiments (data not shown), but this turned out not to be beneficial — probably because adding a regularization term would diminish the contained label information. As both Σ_z and Σ_{avg} require a computation of the epoch-wise covariance $\Sigma(e)$, we propose to choose this matrix as regularization target using the previously mentioned closed-form solution for α^* and ν in order to derive a shrinked estimate $\tilde{\Sigma}(e)$.

3.3 Overview on evaluated SPoC regularization variants

In section 3.1 and 3.2 different regularization strategies were introduced. In Table 3.3 an overview over all proposed approaches is given. The first three rows summarize Tikhonov regularization variants which all require an estimation of α by means of cross-validation. Among them, '**Tik-SPoC**' comprises Tikhonov regularization only according to Eq. 5, while '**NTik-SPoC**' considers an additional trace norm both for $\Sigma(e)$ and Σ_{avg} . The largest extent of regularization is realized by '**ASNTik-SPoC**' which uses the same strategy as 'NTik-SPoC' with additional automatic shrinkage on $\Sigma(e)$ for the computation of Σ_z . As this term applies to the numerator (N) of the objective function, this is marked accordingly in Table 3.3. The last two rows in Table 3.3 summarize automatic shrinkage approaches using the closed-form solution by Ledoit & Wolf (LW). Applying automatic shrinkage to the averaged covariance matrix will be referred to as automatic Tikhonov regularization '**aTik-SPoC**'. In contrast, using automatic shrinkage directly upon

Table 1 Overview on introduced regularization variants for SPoC using two model selection schemes: cross-validation (CV) and based on the Ledoit & Wolf shrinkage estimator (LW). For $\tilde{\Sigma}(e)$ it is explicitly marked if regularization is applied to the numerator (N) and/or to the denominator (D) of the objective function. The checkmarks refer to covariance normalization.

Method	Regularization Target		Normalization
	$\tilde{\Sigma}(e)$	$\tilde{\Sigma}_{avg}$	$\hat{\Sigma}(e), \hat{\Sigma}_{avg}$
Tik-SPoC	-	CV	-
NTik-SPoC	-	CV	✓
ASNTik-SPoC	LW: N	CV	✓
AS-SPoC	LW: N,D	-	-
aTik-SPoC	-	LW	-

$\Sigma(e)$ in the numerator (E) and denominator (D) of the objective function will be referred to as '**AS-SPoC**'.

4 Experiments and validation procedure

4.1 Simulation data

In this work, we aim to characterize and benchmark the introduced regularization techniques to the SPoC $_{\lambda}$ algorithm. However, in the majority of real-world EEG experiments there is no ground truth s available which severely challenges the validation procedure. To compensate for this, a novel data-driven simulation approach for labeling datasets was utilized (Castaño-Candamil et al, 2017). It generates ground-truth label information based on known sources from arbitrary pre-recorded EEG measurements. This post-hoc data labeling allowed obtaining noiseless labels from a relatively large amount of EEG data (here up to 1000 epochs) while conserving the real statistics of the neural activity including non-stationarities of the signal. Furthermore, the approach provided full control over label noise and allows studying its influence upon the decoding performance. In the following, a detailed description for the dataset generation is given.

4.1.1 Preprocessing

In total, 40 datasets of a single motor imagery session per subject formed the basis for the simulation. The experimental design of the motor imagery paradigm is described in detail in (Blankertz et al, 2010). From the recorded EEG, we utilized the signals of 63 passive EEG channels placed according to the extended 10-20 system. The preprocessing of each raw EEG dataset consisted of a high-pass filtering at 0.2 Hz, low-pass filtering at 48 Hz and sub-sampling to 120 Hz. For each dataset, the continuous EEG recordings of active task periods (from the task cue to the end of the imagery interval) were segmented into non-overlapping epochs of 1000 ms duration. Artifact epochs were identified by a min-max threshold and by a variance criterion. The latter was additionally applied to detect and remove

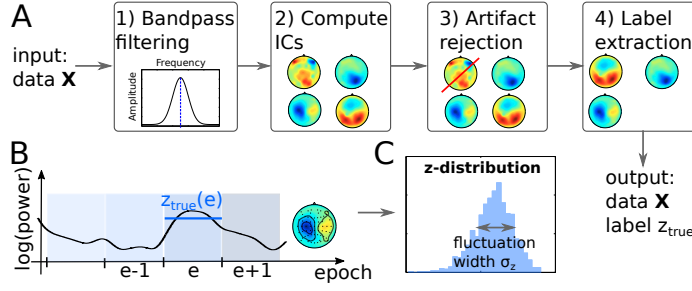


Fig. 1 Procedure for data-driven post-hoc labeling of arbitrary pre-recorded EEG signals. (A) Processing pipeline to extract independent components (ICs). (B) For each IC and epoch e , the log-bandpower average of the epoch serves as a ground truth label $z_{true}(e)$. (C) Distribution of z_{true} over all epochs of an exemplary IC. Its bandpower fluctuation width is described by σ_z .

outlier channels. Details about the artifact preprocessing are described in (Meinel et al, 2016).

As SPoC’s stability depending on the number of training epochs N_e shall be studied in this paper, we discarded datasets with $N_e < 1000$. Similarly, datasets where more than 10% of the original EEG channels had to be rejected, have been removed from further analysis. Applying these criteria, the data of 12 out of 40 subjects remained.

4.1.2 Post-hoc labeling of pre-recorded EEG data

As illustrated in Fig. 1 (A), the following steps were applied to generate continuous labels z_{true} from pre-recorded EEG datasets in a data-driven way:

1. Bandpass filtering of the data to a frequency band of interest. For our analysis, we choose the alpha-band frequency range of $[8, 12]$ Hz.
2. Based on the bandpass filtered data, an ICA decomposition (fastICA, (Hyvarinen, 1999)) into $N_{in} = 20$ independent components (ICs) was computed.
3. To identify and remove artifactual components in an automatic way, the data-driven classification approach MARA (Winkler et al, 2014) for the identification of artifactual components was applied. A posterior probability threshold ($p_{art} = 10^{-8}$) describing the probability of an artifact feature was applied for discarding components of non-neural origin resulting in $N_{sel} \leq N_{in}$ selected ICs.
4. The log-bandpower for each selected component j , with $j = \{1, \dots, N_{sel}\}$, was computed by the Hilbert transform and averaged in each 1 s time interval which defined the epoch-wise known target variable $z_j(e)$ as sketched in Fig. 1 (B).

Overall, the preprocessed data of 12 subjects resulted in 145 oscillatory components (≈ 12 per subject) which survived MARA. For each selected IC, the log-bandpower activation was sampled across $N_e = 1000$ epochs and thus delivered continuous epoch-wise labels $z_{true}(e)$ to the respective epoched EEG signals $\mathbf{X}(e)$.

We expect the SPoC decoding accuracy to be sensitive to the strength of envelope changes of an oscillatory component. The simulation design enables to empirically study this influence by extracting the absolute width in bandpower fluctuations of a single selected IC across a full session. Therefore, we define the

fluctuation width of the j th IC as $\sigma_z := \text{Var}[z_j(e)]$ calculated across the $N_e = 1000$ epochs as illustrated in Fig. 1 (C).

4.1.3 Probing the algorithms under reduced datasets and label noise

In an offline analysis using the generated 145 labeled datasets, all five introduced SPoC regularization variants and the standard SPoC approach were evaluated in a 10-fold chronological CV. For each epoch e , an estimate of the target variable $z_{est}(e)$ was derived according to Eq. 1 by selecting the highest ranked spatial filter obtained from the training data.

To analyze the benefit of regularization under different dataset sizes, we evaluated the algorithms’ stability by systematically reducing each dataset with originally 1000 epochs to smaller data chunks. Therefore, epochs from the session end were removed. For each of the 145 labeled datasets, 22 discrete, logarithmically scaled dataset sizes $N_e \in [20, 1000]$ (respectively training set sizes N_{train}) were tested. Similarly, we probed the stability of our approaches under varying label noise conditions. Therefore, each sample of the target variable distribution z_{true} was modified by adding normally distributed label noise, resulting in a noisy label set z_{noisy} which was used for the CV procedure. According to the label noise model proposed by Castaño et al. (Castaño-Candamil et al, 2017), the correlation between the undistorted and the noisy labels $\rho_n = \text{Corr}(z_{true}, z_{noisy})$ can be controlled via the label noise parameter $\xi_n := 1 - \rho_n$. A value $\xi_n = 1$ refers to maximal label noise, while $\xi_n = 0$ indicates that the labels are completely noise free. Five fixed levels for ξ_n were evaluated.

For the CV-based regularized SPoC variants (see Table 3.3), the regularization strength α was varied in a range $\alpha \in \{0; [10^{-8}, 10^0]\}$. Overall, 20 discrete, logarithmically scaled α levels were analyzed. To summarize, we tested all algorithms on different hyperparameter sets $\Omega = \{(N_{train}, \xi_n, \alpha)\}$.

4.2 Real-world scenario

4.2.1 Dataset for evaluation

In order to examine the regularization methods in a real-world decoding scenario, we utilized data of 18 subjects who participated in a repetitive visuo-motor hand force task with 400 trials per session. The task enables to extract a single-trial motor performance metric such as reaction time (RT) or a cursor path length (CPL). During the full session, EEG from 63 passive Ag/AgCl electrodes (Easy-Cap) placed according to the extended 10-20 system was recorded by multichannel EEG amplifiers (BrainAmp DC, Brain Products) with a sampling rate of 1 kHz. In each trial, a ”get-ready” interval preceded a visually presented ”go-cue”, the latter indicating the start of a motor execution phase. In an offline analysis by Meinel et al (2016), we found that oscillatory bandpower features recorded during the get-ready interval can partially explain upcoming single-trial motor performance. For further details, please see Meinel et al (2016).

Building upon these findings, in this paper the EEG signals were segmented into epochs along the time interval $[-500, +500]$ ms relative to the go-cue in each trial to decode RT of the upcoming motor task. After data preprocessing and

outlier rejection following the workflow described in Meinel et al (2016), we now restricted any further analysis to oscillatory features within the alpha-band frequency range of [8, 12] Hz. The bandpass filter was realized applying a zero-phase butterworth filter of 6th order. The number of epochs N_e surviving the preprocessing varied across subjects from 142 to 352.

4.2.2 Evaluation scheme

All algorithms were evaluated within a (nested) 10-fold chronological CV. The three CV-based regularization variants demanded an additional inner CV to estimate the individually optimal regularization parameter α^* . It was chosen among 15 discrete, logarithmically scaled values in the range $\alpha \in [10^{-8}, 1]$. The α -value maximizing the z-AUC evaluation score (details see section 4.3) was selected and applied to the outer CV in order to train the respective spatial filtering algorithm and linear regression model. The methods 'aTik-SPoC' and 'AS-SPoC' allow for an analytical estimate of α^* and hence did not require an inner CV. In contrast to the simulation scenario, the total number of ground truth neural source(s) which might (partially) explain the target variable z_{true} is not known a priori. By applying a regression model, we assume that several sources might contribute to explain the labels z_{true} .

For each α in the inner or outer CV, the following scheme was applied: a spatial filter set $\{\mathbf{w}^{(i)}\}_{i=1,\dots,N_c}$ was computed on training data \mathbf{x}_{tr} . The first $N_{feat} = 4$ highest ranked components were selected as input to train a linear regression model with coefficients $\{\beta_j\}_{j=0,\dots,N_{feat}}$. The model was trained upon the log-bandpower features $\Phi_{j,tr} = \log(\text{Var}[\mathbf{w}_{tr}^{(j)} \mathbf{x}_{tr}])$. On each feature $\Phi_{j,tr}$, the mean $\mu_{j,tr}$ and the variance $\sigma_{j,tr}$ was estimated in order to standardize the data to zero mean and unit variance before entering the regression model. Given unseen test data \mathbf{x}_{te} , the log-bandpower features $\Phi_{j,te}(e) = \log(\text{Var}[\mathbf{w}_{tr}^{(j)} \mathbf{x}_{te}])$ for each selected spatial filter $\mathbf{w}_{tr}^{(j)}$ were first standardized by $\mu_{j,tr}$ and $\sigma_{j,tr}$. Subsequently, the corresponding coefficients β_j of the trained linear regression model enabled to estimate the target variable $z_{est}(e)$ via:

$$z_{est}(e) = \beta_0 + \sum_{j=1}^{N_{feat}} \beta_j \Phi_{j,te}(e) \quad (10)$$

4.3 Evaluation scores

To compare the estimated labels z_{est} with the known or measured labels z_{true} in the simulation and real-world scenarios across the proposed regularization variants, different evaluation scores can be considered (Meinel et al, 2016). In general, the Pearson correlation coefficient could be utilized but has the drawback, that it is very sensitive to the number of samples (Kenney, 2013). Therefore, we instead decided to utilize the following three scores in this paper:

- **Angle θ between spatial filters:** The design of the simulation scenario gives access to each ground truth spatial filter \mathbf{w}_{true} . As all proposed SPoC variants directly optimize for a spatial filter estimate \mathbf{w} with arbitrary sign and

amplitude (this characteristic is inherited from the formulation as an eigenvalue problem), the angle θ between the spatial filters can directly serve as an evaluation metric:

$$\begin{aligned}\theta_r &= \arccos \left(\frac{\mathbf{w}^\top \mathbf{w}_{true}}{\|\mathbf{w}\| \|\mathbf{w}_{true}\|} \right) \\ \theta &= \begin{cases} \theta_r, & \theta_r \leq \pi/2 \\ \pi - \theta_r, & \theta_r > \pi/2 \end{cases}\end{aligned}\quad (11)$$

with $0 < \theta < \pi$. A perfect decoding will be expressed by an angle $\theta = 0$. Please note that the angle θ can only be estimated within the simulation scenario.

- **Separability z-AUC of labels:** Another possibility is to transfer the continuous labels z_{true} into a two-class scenario according to the median of z_{true} . This enables the utilization of the receiver operating characteristics (ROC) curve which is calculated upon the estimated target variable z_{est} given the true two-class labels (Fawcett, 2006). As ROC performance can be reduced to a scalar value by calculating the area under the ROC curve (AUC), we will name this metric z-AUC as it characterizes the separability of the estimated target variable z_{est} . The z-AUC score can be directly evaluated in both scenarios (Meinel et al, 2017). A perfect decoding corresponds to z-AUC = 1 while chance level correspondents to a value of 0.5.
- **Relative z-AUC performance:** The score z-AUC_{ref} corresponds to the baseline performance of SPoC without any regularization. In this paper, we will compare it to performances obtained by the proposed regularized variants (see Table 3.3). Given a hyperparameter set Ω , the target variable obtained under these hyperparameters $z_{est}(\mathbf{w}(\Omega))$ can be estimated using Eq. 1 and the corresponding z-AUC can be computed. For fixed Ω , the performance of a regularized SPoC variant $\text{z-AUC}_{reg}(\Omega)$ can be assessed as the relative change of z-AUC to the baseline SPoC performance:

$$\text{rel. z-AUC}(\Omega) := \frac{\text{z-AUC}_{reg}(\Omega) - \text{z-AUC}_{ref}(\Omega)}{\text{z-AUC}_{ref}(\Omega)} \quad (12)$$

If $\text{rel. z-AUC} > 0$, this directly corresponds to a relative performance increase compared to SPoC and vice versa.

5 Results

First, we studied the characteristics of the regularization algorithms on 145 analysis problems within the simulation framework. It allows assessing the influence of (hyper)parameters such as regularization strength, dataset size or label noise under controlled conditions. Second, the approaches were tested on real-world data to verify the transferability of the findings and to provide rules of thumb for the practitioner.

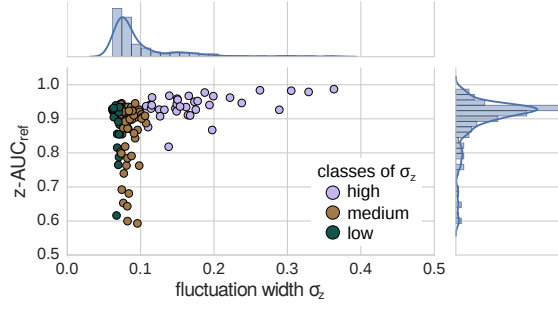


Fig. 2 Simulation data: scatter plot relating the fluctuation width σ_z of each selected ICA component to their baseline SPoC performance $z\text{-AUC}_{ref}$ for non-reduced datasets with 1000 epochs. Based on its σ_z -distribution, the dataset of each IC was labeled into one of three classes, defined by the quartile thresholds Q_{25} and Q_{75} .

5.1 Simulation data

5.1.1 Labeling according to bandpower fluctuation width

The SPoC algorithm optimizes for oscillatory components that co-modulate in their bandpower with a given target variable. In Fig. 2, the relation between the fluctuation width σ_z and the baseline SPoC performance $z\text{-AUC}_{ref}$ on the full dataset $N_e = 1000$ is shown for each of the 145 ICs (correlation $R = 0.31$ with $p = 2.20 \cdot 10^{-4}$). The results indicate that the decoding quality of SPoC depends on the fluctuation width σ_z of the underlying neural component, with stronger fluctuation width being related to higher decoding quality. For further analysis of the simulation data, all 145 ICs were labeled according to their bandpower fluctuation width σ_z into three classes determined by the lower and upper quartile according to the distribution of σ_z across all components (see color coding in Fig. 2). In the following, we will show the decoding performances $z\text{-AUC}_{GA}$ and θ_{GA} as grand average for each corresponding fluctuation width class.

5.1.2 Sensitivity to regularization parameter

Regarding the CV-based regularized SPoC versions, their sensitivity to the regularization parameter α is reported in Fig. 3 exemplarily for the 'high σ_z ' class. It reflects the grand average (GA) of all components contained in this class and provides different evaluation scores. The first row reports the $z\text{-AUC}_{GA}$ while the second row summarizes the angle θ_{GA} between filters. A regularization benefit is expressed via an increasing $z\text{-AUC}_{GA}$ or a decreasing θ_{GA} relative to the performance level at $\alpha = 10^{-8}$. A few observations can be summarized from Fig. 3: First, the two evaluation scores $z\text{-AUC}$ and θ are highly (anti-)correlated across the shown dataset scenarios and SPoC regularization variants. As in real-world data the ground truth will not be known a priori, further analysis will need to be restricted to the metric $z\text{-AUC}$. Second, an increase of the training set size N_{train} (left to right column) leads to a lower sensitivity wrt. α . Third, a comparison of α sensitivity ranges across the three regularization variants yields that 'NTik-SPoC' and 'ASNTik-SPoC' are sensitive in the interval $10^{-6} \leq \alpha \leq 1$ while 'Tik-SPoC' is only sensitive within $10^{-3} \leq \alpha \leq 1$. Fourth, 'NTik-SPoC' and 'ASNTik-SPoC'

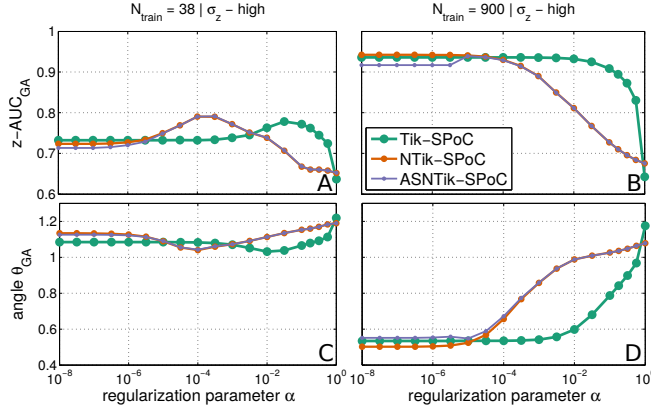


Fig. 3 Simulation results: influence of regularization strength α onto the decoding accuracy of three SPoC variants regularized via CV. The grand average performance $z\text{-AUC}_{GA}$ is reported in the top row (subplots (A) and (B)), while subplots (C) and (D) in the lower row report the angle θ_{GA} between the estimated highest ranked and the ground truth filter as evaluation score. Subplots in the left and right columns differ in the number of training data points (epochs) used for SPoC decoding. Results are reported for the class ' σ_z - high'.

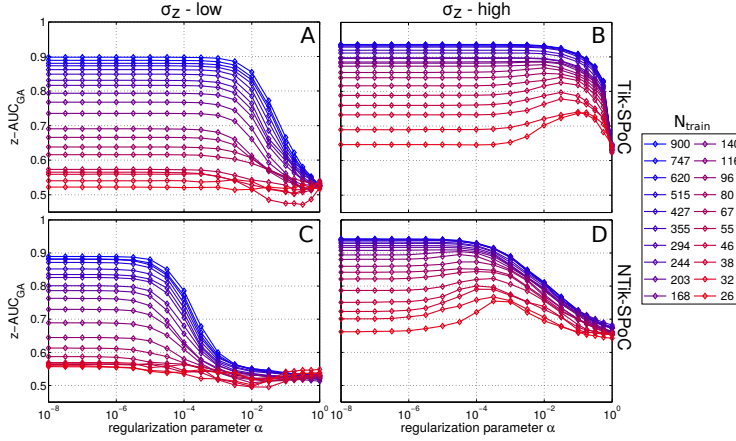


Fig. 4 Simulation results: sensitivity of regularized SPoC variants to α and to reduced training set sizes N_{train} . The grand average performance $z\text{-AUC}_{GA}$ is reported for 'Tik-SPoC' (top row) and 'NTik-SPoC' (bottom row) and separately for the fluctuation width classes 'low' (left column) and 'high' (right column).

behave highly similar, while 'Tik-SPoC' shows a qualitative different behavior. Based on these observations, further analysis will focus on differences between 'NTik-SPoC' and 'Tik-SPoC'. Fifth, extreme regularization with $\alpha = 1$ leads to a drop of decoding performance regardless of the approach, while in the absence of regularization ($\alpha = 0$) a slight improvement due to trace normalization can be reported for 'NTik-SPoC'.

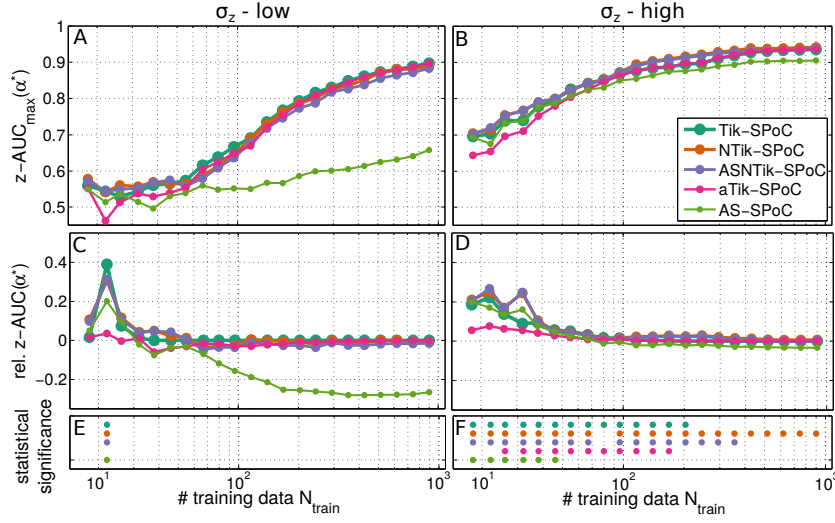


Fig. 5 Simulation results: influence of training set size and fluctuation width upon decoding performance of optimal regularization strength α^* . The top row depicts the grand average **absolute** performance of five regularized SPoC variants for ICs that either have low (A) or high (B) bandpower fluctuation width. The middle row depicts performance increase or decrease of the five regularized methods **relative** to the baseline SPoC method without any regularization and again separately for IC's of low (C) and high (D) fluctuation width. Subplots (E) and (F) reveal color-coded points for each training set size where the regularized variant significantly outperformed the baseline method (Wilcoxon signed rank test with $p < 0.05$).

5.1.3 Influence of reduced datasets and fluctuation width

The simulation scenario grants access to test the stability of different regularized SPoC variants under reduced datasets. For the CV-based methods, a sensitivity analysis for the regularization strength α under 22 training set sizes N_{train} is shown in Fig. 4 for 'Tik-SPoC' (first row) and 'NTik-SPoC' (second row). The two columns in Fig. 4 reveal the influence of the components' fluctuation width σ_z (left: low, right: high). We observed, that regularization has the strongest effects for components with large σ_z and for small training sets. With increasing training set size N_{train} , the sensitivity range for α shifts towards smaller α values. Comparing the depicted methods, 'NTik-SPoC' shows a higher sensitivity to regularization strength α compared to 'Tik-SPoC'. Interestingly, for all subplots (A)-(D) the curves along different N_{train} values converge at $\alpha = 1$, as for this value the SPoC methods collapse to a PCA on the z -weighted covariance. Even for this extreme choice of α , data characterized by higher σ_z reaches a better decoding performance than data with lower σ_z .

To quantify the decoding performances across methods, the maximum GA performance $z\text{-AUC}_{\max} := z\text{-AUC}_{GA}(\alpha^*)$ is reported in Fig. 5 (A) and (B) in the absence of label noise. Therefore, the optimal regularization strength $\alpha^* = \arg \max_{\alpha} z\text{-AUC}_{GA}(\alpha)$ is selected for fixed N_{train} and σ_z class. For variants using the LW estimate, this selection is not necessary as there is an analytic solution for α^* such that $z\text{-AUC}_{\max} = z\text{-AUC}_{GA}$. Accordingly, the relative performance change $\text{rel. } z\text{-AUC}(\alpha^*)$ is reported on the GA level in Fig. 5 (C) and (D), while (E)

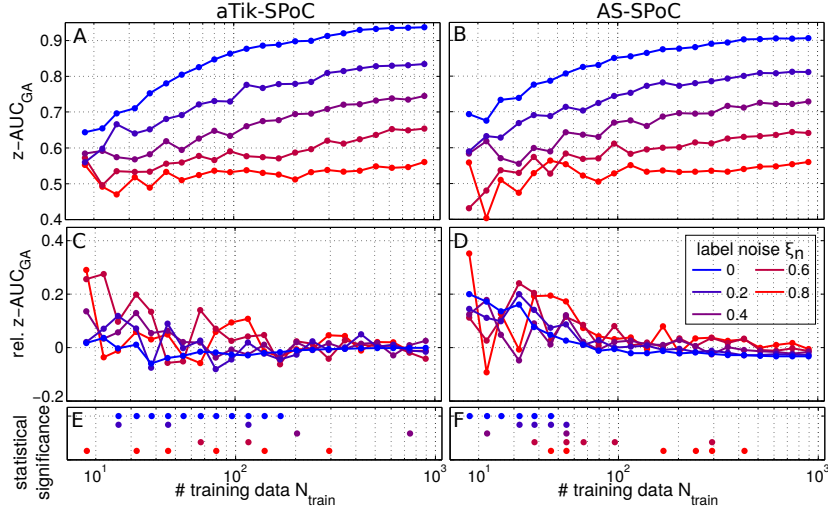


Fig. 6 Simulation results: interaction between label noise level ξ_n and dataset size N_{train} . A level of $\xi_n = 0$ states the absence of label noise. All curves report the grand average results for ICs belonging to the 'high σ_z ' class. Subplots (A) and (B) in the top row provide the **absolute** grand average performances for 'aTik-SPoC' and 'AS-SPoC', while the middle row depicts **relative** performance changes. The dots in (E) and (F) indicate configurations, for which the regularized variant significantly outperformed the baseline method (Wilcoxon signed rank test with $p < 0.05$).

and (F) report the statistical significance of the findings. Therefore, a one-sided Wilcoxon signed rank test was applied to test if the median of performance differences ($z\text{-AUC}_{max,ref}(\Omega) - z\text{-AUC}_{max,reg}(\Omega)$) is smaller or equal to zero for fixed N_{train} and σ_z . If a p-value $p < 0.05$ was found (not corrected for multiple testing), the configuration Ω reveals a significant difference among the two methods, indicated by a colored data point in Fig. 5 (E) and (F). The following observations can be reported: First, the absolute decoding performance strongly depends on N_{train} regardless of the regularization method and σ_z class. Second, there is a relative performance increase of all introduced regularization methods up to training sets of size $N_{train} \approx 60$ on the grand average level. For larger datasets the regularization does not reveal an additional benefit on the grand average. Third, our results indicate that regularization is beneficial for various methods in the 'high σ_z ' class, while this is not the case for 'low σ_z '. Here, a noticeable case is reported by the performance of 'AS-SPoC' which drastically loses performance for $N_{train} \gtrsim 50$.

5.1.4 Stability under label noise and reduced data

As in most real-world scenarios label noise challenges the decoding performance of subspace methods like SPoC (Castaño-Candamil et al, 2015). Thus, we studied its influence for reduced datasets within the simulation data. Fig. 6 exemplary shows the degrading decoding performance under label noise conditions for 'aTik-SPoC' and 'AS-SPoC' for 'high σ_z '. Both methods have in common, that performance estimates are very noisy under small dataset size and increasing label noise. Regarding the maximally achievable decoding performance for both meth-

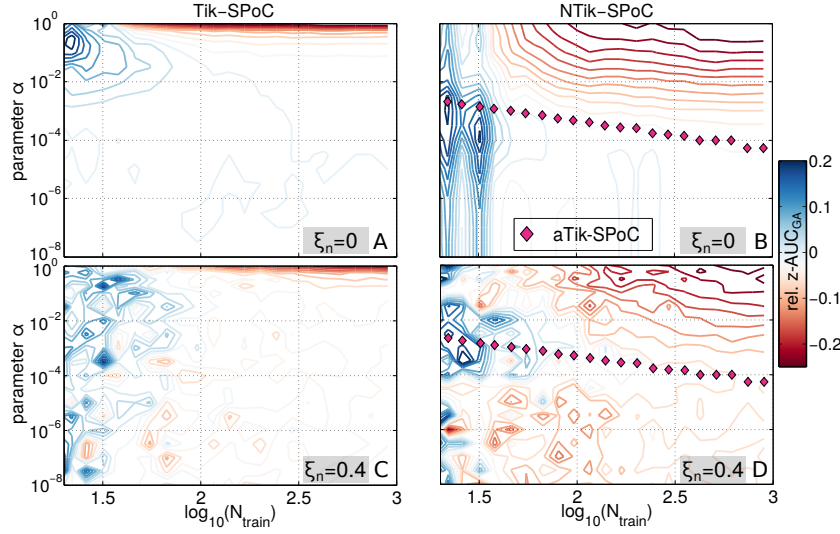


Fig. 7 Simulation results: landscape of the grand average **relative** performance changes in $z\text{-AUC}_{GA}$ dependent on the training set size N_{train} and regularization strength α for ICs of high fluctuation width σ_z . The isolines of relative performance changes were interpolated along a grid search. No label noise was applied to generate maps (A) and (B) for methods 'Tik-SPoC' and 'NTik-SPoC', respectively. The second row reports the landscapes including a label noise level of $\xi_n = 0.4$ for both methods. Additional diamond markers in subplots (B) and (D) depict the grand average of α^* for 'aTik-SPoC', which is independent of label noise. This method utilizes analytically derived values of α^* and may serve as a reference for the CV-based 'NTik-SPoC'.

ods at $N_{train} = 900$, the absolute performance $z\text{-AUC}_{GA}$ scales almost linearly with the amount of label noise ξ_n . Referring to the relative performance change shown in (C) and (D) as well as the statistical tests in (E) and (F), they reveal that under increased levels of label noise ξ_n even larger training sets can profit from regularization when compared to the unregularized SPoC. While for $\xi_n = 0$ a relative performance increase on the GA can be found up to $N_{train} \approx 60$, for $\xi_n = 0.6$ it increases up to $N_{train} \approx 800$. This effect is stronger for 'AS-SPoC' than for 'aTik-SPoC'. Despite not shown here, we would like to mention, that under increased label noise the performance gain of the regularized variants with larger N_{train} can be observed also for the 'low σ_z ' case, but with a lower overall decoding performance.

5.1.5 Optimal regularization parameter ranges

To identify suitable ranges of the regularization parameter for the CV-based methods, color-coded contour maps of relative performance changes are provided in Fig. 7. The maps show the grand average $\text{rel. } z\text{-AUC}_{GA}$ within the (N_{train}, α) hyperparameter space separately for the two methods 'Tik-SPoC' (first column) and 'NTik-SPoC' (second column). Maps in the upper row summarize the performance changes in the absence of label noise ($\xi_n = 0$) while the lower one provides these results under systematic label noise ($\xi_n = 0.4$). The blue areas in each map mark ranges in the hyperparameter space, where a relative performance increase is

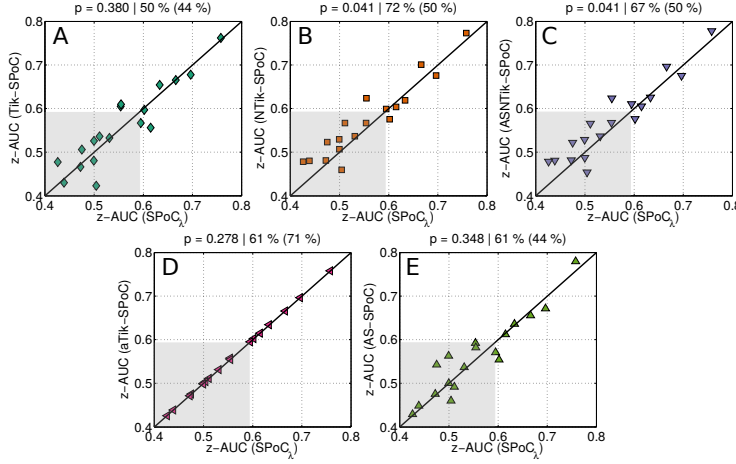


Fig. 8 Real data: scatter plots (A)-(E) compare the performance of different regularized SPoC variants with the unregularized baseline method SPoC. In each subplot, a marker represents one of 18 subjects. Above each scatter plot, the p-value of a one-sided Wilcoxon rank sum test is given as well as the percentage of subjects for which the regularized variant outperforms baseline SPoC. Additional percentage values in brackets exclude data points located inside the grey shaded area. The latter marks a threshold criterion on z-AUC for meaningful predictions.

obtained, while "no-go" areas in red associate with a decrease of decoding quality. When comparing Fig. 7 (A) and (B), we observe that the trace norm in 'NTik-SPoC' induces a reduction of optimal α values by a few orders of magnitude as well as a larger sensitivity range compared to 'Tik-SPoC'. Both plots reveal consistently a "no-go" area towards the top right corner, which indicates, that on the grand average strong regularization is detrimental, when large training datasets without label noise are available. With additional label noise in Fig. 7 (C) and (D), the heterogeneity of the relative performance landscape increases and the "no-go" areas at the top right shift towards larger N_{train} . In accordance with the automatic shrinkage based methods visualized in Fig. 6 we find, that the inclusion of label noise ξ_n into the simulation has the effect that regularization might even be beneficial for large training sets.

For different training set sizes N_{train} , we now compare the CV-based estimates of α with those of 'aTik-SPoC', which makes use of an analytical solution α^* . The grand average of α^* is plotted into Fig. 7 (B) and (D). As the analytical solution for α^* (Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005) is proportional to N_{train}^{-2} , it should scale anti-proportional with $\log_{10}(N_{train})$, which in fact was observed in Fig. 7 (B). It is worth to mention, that the analytic choices of α^* are not influenced by label noise – compare maps (B) and (D) – as the involved covariance shrinkage (see Eq. 9) does not make use of the label information.

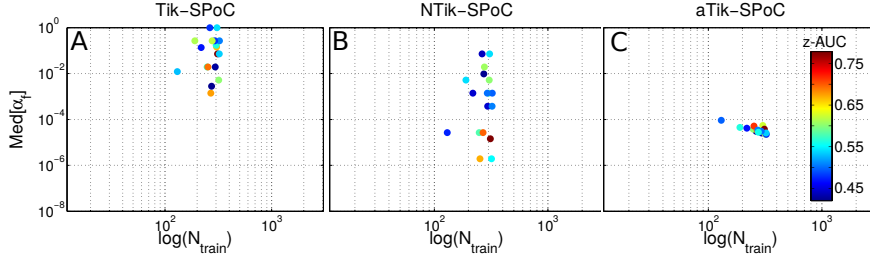


Fig. 9 Real data: median regularization strength across the 10-fold chronological cross-validation for each dataset as a function of the training set size N_{train} color coded by the achieved z-AUC decoding performance.

5.2 Real-world data scenario

5.2.1 Comparison of regularized SPoC variants

In Fig. 8, the subject-wise performance comparison of all regularized SPoC variants to standard SPoC is depicted. To compare each regularized variant to its baseline, we report two different group statistics. First, the overall ratio of subjects for which the regularized variant outperforms standard SPoC is provided. Second, the values in brackets consider only those individual performances which cross a threshold of minimum meaningful performance $z\text{-AUC}_{th} = 0.59$. For details on how this chance level has been determined via group analysis of predictors, we refer to Meinel et al (2016). To verify if a regularized variant reaches a statistically significantly higher performance compared to standard SPoC, a one-sided Wilcoxon rank sum test was evaluated on the group level. The corresponding p-values are reported in the plot headers of Fig. 8 (A)-(E).

The following observations for the RT decoding on real-world data were made: First, in contrast to all other regularization variants, the performance changes induced by 'aTik-SPoC' are negligible small. Second, across the remaining regularization approaches we observed a tendency towards larger benefits for initially poorly performing subjects. On the group level, all regularization methods except 'Tik-SPoC' registered the majority of data points above the bisectrix. The CV-based 'NTik-SPoC' and 'ASNTik-SPoC' behave very similarly, which have been observed before on simulation data (see Fig. 5). Both approaches significantly outperform the baseline SPoC performance.

5.2.2 Selected regularization strengths

The regularization parameter values α_f obtained on real data by the nested CV-based regularization variants across folds f are evaluated in Fig. 9 (A) and (B). Its plots should be compared with the maps for simulations depicted in Fig. 7. The median $\text{Med}[\alpha_f]$ across folds is shown for each subject. It's color encodes the associated z-AUC performance. The results indicate that 'NTik-SPoC' operates in smaller α ranges than 'Tik-SPoC' does, which is in accordance with the observations from the simulation in Fig. 7. For the majority of subjects, the regularization strength is outside the "no-go" areas of the simulation as α was selected by nested CV from the interval $[10^{-8}, 1]$. For a few subjects, a large α was chosen.

As expected from simulations, this strong regularization is linked with a low absolute decoding level. The median of the analytically computed α_f^* across folds for 'aTik-SPoC' are presented in Fig. 9 (C). For most subjects a way smaller median regularization strength is chosen compared to the CV-based 'NTik-SPoC' method, while we observe that the analytical solution does not elicit a significant decoding improvement (see Fig. 8 (D)).

6 Discussion

In summary, we have proposed a set of novel regularization techniques for SPoC. We investigated their effectiveness by evaluating their performance both on simulated and on real-world datasets. Overall, 'NTik-SPoC' based on Tikhonov regularization and additional covariance normalization turned out to be the most beneficial technique.

Simulation scenario

A closer look upon the simulation results clearly shows that the regularization benefit for SPoC strongly depends on the dataset size, prevalent label noise conditions as well as on the fluctuation width of the underlying component. As a strong absolute performance variability across datasets was present in the simulation, the reported grand average performance provides a way less optimistic view than single dataset results do. Largest regularization benefit was reported for low amount of data and components with large fluctuation widths. The latter observation might be explained by the intrinsic difficulty of SPoC to recover sources of small bandpower changes.

Intuitively, additional label noise reduces the information content per data point such that the estimation of Σ_z gets more demanding. Theoretically, this disadvantage could be compensated by either enlarging the training set or by adding regularization. Using the large amount of simulation data, we were able to show that under label noise conditions even larger datasets profit from regularization.

Surprisingly, in the simulation we found that 'AS-SPoC' loses performance for large datasets (especially for 'low σ_z ' components) while it outperformed standard SPoC on small datasets and revealed a good performance on real-world data as well. This observation might be explained as follows: In the simulation data, the target variable is directly estimated from the EEG (IC) epoch. As such, there should be enough samples in each epoch to estimate reliably the target variable, since it was created this way. Epoch regularization might thus not be necessary here. However, for real data this might not be the case, as the target variable is not directly dependent on the EEG epoch and contains an even unknown label noise level. As such, epoch regularization might be much more useful in that case.

The direct transferability of the simulation results to real-world data is limited by three major differences: First, in real-world experiments the number of neural sources is not known a priori. Thus, a good decoding of source power typically requires the use of several components and of a regression model. Second, in real-world experiments both, label noise and the components' fluctuation widths act as latent variables and cannot directly be estimated. Third, while in the simulation we can almost perfectly recover the label information given sufficient amount of

data ($z\text{-AUC} > 0.9$), in real-world experiments we clearly expect a decreased upper limit of the decoding performance. This is a strong indicator for the assumption, that solely bandpower information may not suffice to fully explain the labels.

Real-world scenario

Based on the real-world data, we could show that predominantly the decoding performance of initially poorly performing subjects was improved by almost all regularized SPoC approaches (except 'aTik-SPoC'). However, we cannot report a single regularization variant that systematically performed best on all subjects.

Two important aspects can be transferred from the simulation to the real-world data. First, the simulation allowed deriving an operating range of the regularization hyperparameter α for each CV-based regularization variant. When comparing these findings with the real-world data, we found that the optimal choice of the regularization intensity α for the CV-based techniques is in good accordance with the derived "no-go" areas obtained from our simulations.

Second, according to the simulation under label noise in Fig. 6, we could gain an estimate of the label noise conditions ξ_n of any real-world dataset directly by comparing the absolute achievable decoding levels with the real-world decoding performances in Fig. 8. As an example, for the best performing subject of Fig. 8 (E) with $z\text{-AUC} \approx 0.78$ on $N_{train} = 310$ data points, the label noise level can be estimated as $\xi_n \approx 0.2$ according to Fig. 6 (B). Despite such estimates may not perfectly represent the ground truth, they might be beneficial for comparing data from multiple experimental paradigms e.g. in order to choose most suitable regularization strategies.

CV-based vs. analytical model selection

Overall, we introduced three CV-based Tikhonov regularization methods for SPoC (see overview in Table 3.3) and compared their performance against two variants based on automatic covariance shrinkage. Although the decoding performance of all three Tikhonov variants are on comparable levels, they strongly differ in terms of their sensitivity range for the regularization parameter. This information, however, is of great importance when it comes to choosing parameters by cross-validation. Interestingly, we found that 'NTik-SPoC' and 'ASNTik-SPoC' profit from a logarithmically scaled search space wrt. regularization parameter α while 'Tik-SPoC' could also cope with a linear scaling. We conclude that this behavior is introduced by the additional trace normalization. When comparing 'NTik-SPoC' and 'ASNTik-SPoC', the inclusion of additional LW-based shrinkage for the numerator regularization realized by 'ASNTik-SPoC' does not boost performance significantly. Accordingly, 'NTik-SPoC' seems preferable in a direct comparison due to its lower computational effort. In future work, an alternative data-driven estimation of the regularization parameter without cross-validation might be achieved e.g. by utilizing a Bayesian framework which estimates the regularization strength via expectation maximization (Mattout et al, 2006).

Comparing both LW-based covariance shrinkage based approaches, 'AS-SPoC' seems to be the better choice compared to 'aTik-SPoC'. Three arguments support this view. First, referring to the label noise challenged simulation in Fig. 6 we found that 'AS-SPoC' profits from regularization under high label noise even for

larger training sets ($N_{train} \gtrsim 300$) while this effect was less pronounced for 'aTik-SPoC'. Second, we found that the analytically derived regularization parameter for 'aTik-SPoC' across subjects is chosen way smaller compared to values chosen by CV for 'NTik-SPoC'. For 'aTik-SPoC', the concatenation of epochs results in $N_s \cdot N_{train}$ sample points to estimate Σ_{avg} . As the LW-based regularization parameter is anti-proportional to the number of samples (Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005), an overly small regularization parameter is chosen, irrespectively of whether the covariance estimate did improve. Third, the analytic approach makes an i.i.d. assumption about the data. A violation thereof due to outliers might be compensated with a CV-based strategy but not by 'aTik-SPoC'. The i.i.d. assumption might also be violated for 'AS-SPoC' when the LW-based analytical solution for the trial-wise covariance matrix is challenged by autocorrelated data of a single epoch. A potential mitigation may be provided by alternative covariance shrinkage estimators that accounts for autocorrelated data as proposed by Bartz et al. (Bartz and Müller, 2014). Alternatively closed-form solutions for covariance shrinkage assuming elliptical distributions could also prove superior to the LW-based solution (Chen et al, 2011).

Guidance for the practitioner

Both, simulation and real-world data results strongly indicate that there is not one single regularization variant that outperforms all others. Different global parameters, such as dataset size, the noise conditions or non-stationarity in the data influence the achievable decoding accuracy.

The work by Engemann and Gramfort (2015) reported the superiority of CV-based compared to analytical model selection in the context of spatial whitening of M/EEG data. This supports our proposal to prefer the CV-based approaches 'Tik-SPoC' or 'NTik-SPoC' over the LW-based 'AS-SPoC' method. All three methods, however, are analytically solvable by an eigenvalue decomposition and require relatively low computational effort. As they may come up with partially disjunct components, we thus propose in practice to evaluate all three variants in parallel. The final feature set should be selected in a data-driven strategy to deduce the overall most relevant oscillatory components for a given application scenario.

7 Conclusion

We investigated novel regularization variants for SPoC and reported their characteristics in a simulation and real-world data scenario. Initially, we applied a novel data-driven simulation framework that by design enables to generate labeled EEG datasets. The simulation delivered two main results:

First, it allowed comparing and explaining characteristics of the regularized SPoC algorithms. We could study the influence of varying training set sizes, label noise and of the bandpower fluctuation width of the neural sources of interest. On the one hand, we found that the achievable overall decoding performance decays under increased label noise conditions and smaller datasets. On the other hand, small datasets and label noise were the settings under which several regularized SPoC variants could outperform the original unregularized algorithm. As most real-world experiments come with an unknown amount of label noise, we expect

that the benefits of regularization would transfer into real-world problems. Second, the simulation outcomes offered a guideline for practitioners. It proposes to tune the search for a suitable regularization parameter to a log-scaled search space. Furthermore, it indicates that the number of training data points and label noise present in the data should guide the choice of this parameter.

As an additional validation, we tested the regularized SPoC algorithms on real-world EEG data. Its outcome supported the guidelines obtained by simulation concerning the choice of regularization parameters and achievable performance improvements. We found, that individual datasets could profit strongly from single forms of regularization. As a consequence, we recommend testing several versions of regularization if decoding performance is to be optimized in practice.

While we have chosen to compare relatively simple and general regularization techniques, this work could be expanded to more sophisticated regularization strategies e.g. to realize session-to-session or subject-to-subject transfer scenarios. The presented regularization framework and the evaluation strategy using simulated and real-world datasets may pave this way.

Information Sharing Statement

The Matlab code for the proposed SPoC regularizations is accessible on GitHub under https://github.com/ameinel/regularized_SPoC. The datasets are available upon request from the authors.

Acknowledgements This work was fully supported by BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG), grant number EXC 1086. For the data analysis, the authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) by grant no. INST 39/963-1 FUGG. Fabien Lotte received research support from the French National Research Agency with the REBEL project (grant ANR-15-CE23-0013-01) and the European Research Council with the BrainConquest project (grant ERC-2016-STG-714567). For parts of the data analysis, the Matlab-based BBCI toolbox was utilized (Blankertz et al, 2016). The authors declare that they have no conflict of interest. This is a post-peer-review, pre-copyedit version of an article published in *Neuroinformatics* (Springer). The final authenticated version is available online at: <http://dx.doi.org/10.1007/s12021-018-9396-7>

References

- Arvaneh M, Guan C, Ang KK, Quek C (2011) Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering* 58(6):1865–1873, DOI 10.1109/TBME.2011.2131142
- Arvaneh M, Guan C, Ang KK, Quek C (2013) Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems* 24(4):610–619, DOI 10.1109/TNNLS.2013.2239310
- Bai Z, Silverstein JW (2009) *Spectral Analysis of Large Dimensional Random Matrices*. Springer Science & Business Media
- Bartz D, Müller KR (2014) Covariance shrinkage for autocorrelated data. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) *Ad-*

- vances in Neural Information Processing Systems 27, Curran Associates, Inc., pp 1592–1600
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. *Signal Processing Magazine, IEEE* 25(1):41–56
- Blankertz B, Sannelli C, Halder S, Hammer EM, Kübler A, Müller KR, Curio G, Dickhaus T (2010) Neurophysiological predictor of SMR-based BCI performance. *NeuroImage* 51(4):1303–1309, DOI 10.1016/j.neuroimage.2010.03.022
- Blankertz B, Acqualagna L, Dähne S, Haufe S, Schultze-Kraft M, Sturm I, Uuml M, Wenzel MA, Curio G, Müller KR (2016) The berlin brain-computer interface: Progress beyond communication and control. *Frontiers in Neuroscience* 10:530, DOI 10.3389/fnins.2016.00530
- Castañó-Candamil JS, Meinel A, Dähne S, Tangermann M (2015) Probing meaningfulness of oscillatory EEG components with bootstrapping, label noise and reduced training sets. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE*, pp 5159–5162
- Castañó-Candamil S, Meinel A, Tangermann M (2017) Post-hoc labeling of arbitrary EEG recordings for data-efficient evaluation of neural decoding methods. *arXiv preprint arXiv:171108208*
- Chen Y, Wiesel A, Hero AO (2011) Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing* 59(9):4097–4107, DOI 10.1109/TSP.2011.2138698
- Cheng M, Lu Z, Wang H (2017) Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cognitive Neurodynamics* 11(2):173–181, DOI 10.1007/s11571-016-9417-x
- de Cheveigné A, Parra LC (2014) Joint decorrelation, a versatile tool for multi-channel data analysis. *NeuroImage* 98(Supplement C):487–505, DOI 10.1016/j.neuroimage.2014.05.068
- Cho H, Ahn M, Kim K, Jun SC (2015) Increasing session-to-session transfer in a braincomputer interface with on-site background noise acquisition. *Journal of Neural Engineering* 12(6):066,009, DOI 10.1088/1741-2560/12/6/066009
- Clerc M, Bougrain L, Lotte F (2016) *Brain-Computer Interfaces 2: Technology and Applications*. John Wiley & Sons
- Dähne S, Meinecke FC, Haufe S, Höhne J, Tangermann M, Müller KR, Nikulin VV (2014) SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage* 86(0):111–122, DOI <http://dx.doi.org/10.1016/j.neuroimage.2013.07.079>
- De Bie T, Cristianini N, Rosipal R (2005) Eigenproblems in pattern recognition. In: *Handbook of Geometric Computing*, Springer, pp 129–167
- De Vos M, Riès S, Vanderperren K, Vanrumste B, Alario FX, Huffel VS, Burle B (2010) Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 8(2):135–150, DOI 10.1007/s12021-010-9071-0
- Devlaminc D, Wyns B, Grosse-Wentrup M, Otte G, Santens P (2011) Multisubject learning for common spatial patterns in motor-imagery BCI. *Intell Neuroscience* 2011:8:8–8:8, DOI 10.1155/2011/217987
- Engemann DA, Gramfort A (2015) Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. *NeuroImage* 108:328–342, DOI 10.1016/j.neuroimage.2014.12.040

- Farina D, Jiang N, Rehbaum H, Holobar A, Graimann B, Dietl H, Aszmann OC (2014) The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22(4):797–809, DOI 10.1109/TNSRE.2014.2305111
- Farquhar J, Hill NJ (2013) Interactions between pre-processing and classification methods for event-related-potential classification. *Neuroinformatics* 11(2):175–192, DOI 10.1007/s12021-012-9171-0
- Farquhar J, Hill N, Lal, TN, Schölkopf, B (2006) Regularised CSP for sensor selection in BCI. In: *Proceedings of the 3rd international BCI workshop*
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Letters* 27(8):861–874
- Frey J, Daniel M, Hachet M, Castet J, Lotte F (2016) Framework for electroencephalography-based evaluation of user experience. In: *Proceedings of CHI*, pp 2283–2294
- Grosse-Wentrup M, Liefhold C, Gramann K, Buss M (2009) Beamforming in non invasive brain-computer interfaces. *IEEE Transactions on Biomedical Engineering* 56(4):1209 – 1219
- Hyvarinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10(3):626–634, DOI 10.1109/72.761722
- Kang H, Nam Y, Choi S (2009) Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters* 16(8):683–686, DOI 10.1109/LSP.2009.2022557
- Kenney JF (2013) *Mathematics of statistics*. D. Van Nostrand Company Inc; Toronto; Princeton; New Jersey; London; New York; Affiliated East-West Press Pvt-Ltd; New Delhi
- Koles ZJ (1991) The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and clinical Neurophysiology* 79(6):440–447
- Krusienski D, Grosse-Wentrup M, Galán F, Coyle D, Miller K, Forney E, Anderson C (2011) Critical issues in state-of-the-art brain-computer interface signal processing. *Journal of Neural Engineering* 8(2):025,002
- Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2):365–411, DOI 10.1016/S0047-259X(03)00096-4
- Lotte F (2015) Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. *Proceedings of the IEEE* 103(6):871–890, DOI 10.1109/JPROC.2015.2404941
- Lotte F, Guan C (2010) Learning from other subjects helps reducing brain-computer interface calibration time. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp 614–617, DOI 10.1109/ICASSP.2010.5495183
- Lotte F, Guan C (2011) Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering* 58(2):355–362, DOI 10.1109/TBME.2010.2082539
- Lu H, Eng HL, Guan C, Plataniotis KN, Venetsanopoulos AN (2010) Regularized common spatial pattern with aggregation for EEG classification in small-sample setting. *IEEE Transactions on Biomedical Engineering* 57(12):2936–2946, DOI

- 10.1109/TBME.2010.2082540
- Mahmud M, Kaiser MS, Hussain A, Vassanelli S (2018) Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Transactions on Neural Networks and Learning Systems* 29(6):2063–2079, DOI 10.1109/TNNLS.2018.2790388
- Makeig S, Debener S, Onton J, Delorme A (2004) Mining event-related brain dynamics. *Trends in Cognitive Sciences* 8(5):204 – 210, DOI <https://doi.org/10.1016/j.tics.2004.03.008>
- Makeig S, Kothe C, Mullen T, Bigdely-Shamlo N, Zhang Z, Kreutz-Delgado K (2012) Evolving signal processing for brain-computer interfaces. *Proceedings of the IEEE 100(Special Centennial Issue):1567–1584*, DOI 10.1109/JPROC.2012.2185009
- Mattout J, Phillips C, Penny WD, Rugg MD, Friston KJ (2006) MEG source localization under multiple constraints: An extended Bayesian framework. *NeuroImage* 30(3):753–767, DOI 10.1016/j.neuroimage.2005.10.037
- Meinel A, Castaño-Candamil S, Reis J, Tangermann M (2016) Pre-trial EEG-based single-trial motor performance prediction to enhance neuroergonomics for a hand force task. *Frontiers in Human Neuroscience* 10:170, DOI 10.3389/fnhum.2016.00170
- Meinel A, Lotte F, Tangermann M (2017) Tikhonov regularization enhances EEG-based spatial filtering for single-trial regression. In: *Proceedings of the 7th Graz Brain-Computer interface Conference 2017*, pp 308–313, DOI 10.3217/978-3-85125-533-1-57
- Millán JdR, Rupp R, Mueller-Putz G, Murray-Smith R, Giugliemma C, Tangermann M, Vidaurre C, Cincotti F, Kübler A, Leeb R, Neuper C, Müller KR, Mattia D (2010) Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges. *Frontiers in Neuroscience* 4:161, DOI 10.3389/fnins.2010.00161
- Nicolae IE, Acqualagna L, Blankertz B (2017) Assessing the depth of cognitive processing as the basis for potential user-state adaptation. *Frontiers in Neuroscience* 11, DOI 10.3389/fnins.2017.00548
- Park SH, Lee D, Lee SG (2017) Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP(99):1–1, DOI 10.1109/TNSRE.2017.2757519
- Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. *NeuroImage* 28(2):326–341
- Ramanathan C, Ghanem RN, Jia P, Ryu K, Rudy Y (2004) Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. *Nature Medicine* 10(4):nm1011, DOI 10.1038/nm1011
- Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. *Rehabilitation Engineering, IEEE Transactions on* 8(4):441–446
- Reuderink B, Poel M (2008) Robustness of the common spatial patterns algorithm in the BCI-pipeline. Tech. rep., HMI, University of Twente
- Samek W, Vidaurre C, Müller KR, Kawanabe M (2012) Stationary common spatial patterns for brain-computer interfacing. *Journal of Neural Engineering* 9(2):026,013, DOI 10.1088/1741-2560/9/2/026013

- Samek W, Meinecke FC, Müller KR (2013) Transferring subspaces between subjects in brain-computer interfacing. *IEEE Transactions on Biomedical Engineering* 60(8):2289–2298, DOI 10.1109/TBME.2013.2253608
- Samek W, Kawanabe M, Müller KR (2014) Divergence-based framework for common spatial patterns algorithms. *IEEE Reviews in Biomedical Engineering* 7:50–72, DOI 10.1109/RBME.2013.2290621
- Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* 4(1)
- Schultze-Kraft M, Dähne S, Gugler M, Curio G, Blankertz B (2016) Unsupervised classification of operator workload from brain signals. *Journal of Neural Engineering* 13(3):036,008, DOI 10.1088/1741-2560/13/3/036008
- Tian TS, Huang JZ, Shen H, Li Z (2013) EEG/MEG source reconstruction with spatial-temporal two-way regularized regression. *Neuroinformatics* 11(4):477–493, DOI 10.1007/s12021-013-9193-2
- Tikhonov AN (1963) Regularization of incorrectly posed problems. *Soviet Mathematics Doklady* 4:1624–1627
- Úbeda A, Azorn JM, Chavarriaga R, R Millán Jd (2017) Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques. *Journal of NeuroEngineering and Rehabilitation* 14:9, DOI 10.1186/s12984-017-0219-0
- Wang H, Li X (2016) Regularized filters for L1-norm-based common spatial patterns. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24(2):201–211, DOI 10.1109/TNSRE.2015.2474141
- Winkler I, Brandl S, Horn F, Waldburger E, Allefeld C, Tangermann M (2014) Robust artifactual independent component classification for bci practitioners. *Journal of Neural Engineering* 11(3):035,013
- Wu D, King JT, Chuang CH, Lin CT, Jung TP (2017) Spatial filtering for EEG-based regression problems in brain-computer interface (BCI). *IEEE Transactions on Fuzzy Systems* PP(99):1–1, DOI 10.1109/TFUZZ.2017.2688423